



Universidad  
de Navarra

**XLII Seminario Interuniversitario de Teoría de la Educación**

## **Hacia una teoría de la educación transformadora**

Pamplona, del 19 al 21 de mayo de 2024  
Universidad de Navarra

### **LA IA y su poder transformador: desafíos e interrogantes**

(ponencia 3)

Isabel Álvarez Cánovas

*Universidad Autónoma de Barcelona*

Antes de empezar, quisiera poder felicitar a los autores de esta ponencia por ofrecernos un texto que nos invita a la reflexión e instiga a proponer diversos puntos para su debate. Además, su presencia en este Seminario de Teoría de la Educación nos propone el disfrute intelectual de poder contraponer y sustentar un espacio vital para poder encajar los cambios profundos y las decisiones en nuestro día a día como docentes. En este sentido, esta agenda pivotará sobre la noción de proponer algunos argumentos sobre la idea de la educación transformadora mediada por la perspectiva casi omnipresente de la IA regenerativa y máxime cuando nos referimos al uso de los *Large Language Models* (LLMs) en sus múltiples caras y formas.

Es notorio que la irrupción de los LLMS, en concreto, entre los modelos más conocidos y ampliamente utilizados como el ChatGPT o Google Bard nos están imprimiendo una profunda reestructuración del propio término de *educación transformadora* que teníamos asumido hasta el momento cuando, por ejemplo, se nos eleva a la noción de propia superación de la condición humana, es decir, de una nueva reconceptualización transhumanista. En este sentido la educación transhumanista se presenta mediante la IA como un nuevo *totalitarismo*, y donde se impone un cierto uso del terror, como bien sustenta Arendt, (1951; 2017)

Terror as the execution of a law of movement whose ultimate goal is not the welfare of men or the interest of one man but the fabrication of mankind, eliminates individuals for the sake of the species, sacrifices the 'parts' for the sake of the 'whole.' (p. 496)

Creo que esta superación transhumanista no se ha presentado en toda su capacidad hasta ahora, posiblemente por esta preocupación de superación, no antes concebida en lo que se refiere a la máquina sobre la persona, aunque sí que hemos tenido ya diversos apuntes de esta

superación antropomórfica (agentes artificiales, Lee 2024) y de sus consecuencias para con los humanos tanto en la literatura, con la obra de la novelista inglesa Mary W. Shelly y su Frankenstein (1816), la ya clásica cinta de Lang (1927), Metrópolis, que curiosamente se desarrolla en un futuro a cien años vista, situándose en el año 2026, es decir solo dos años respecto con el año actual, donde la persona, el sujeto creador se ve superado por su propia producción. Por otro lado, Capek (1920), también incide en la confrontación entre robots (o llamados hombres artificiales) y humanos en su obra teatral *Robots Universales Rossum*. Cuando esta superación de la calidad humana se sucede, entonces entramos ya en modo de amenaza como especie, es decir, en el *déjà vu* del ideario colectivo sobre la selección Natural de las especies Darwiniano, pero esta vez delante una especie para la que no nos habíamos anticipado.

Hasta ahora, a pesar de las primeras cautelas y soslayando futuras predisposiciones, esta superación no suponía una amenaza en términos inminentes, ya que el control humano, podía ser usado en caso de necesidad, es decir, todo continuaba controlado porque la máquina se circunscribía a una mera acción ejecutora de nuestras órdenes y/o necesidades, mediante instrucciones más o menos complejas de codificación. O digámoslo así, la amenaza se podría reducir más, a título individual que el de un sentir comunitario, pero todo esto empezó a cambiar cuando, se fueron sucediendo informaciones recientes y desde dentro de los círculos de los profesionales de la ingeniería de la IA. Así, en marzo del año 2023, un grupo de expertos publicaron una carta abierta para pedir una pausa de seis meses y así detener algunos aspectos del desarrollo en la investigación sobre la IA, pausa que escribiendo estas líneas ya se ha sobrepasado y no disponemos de los logros que hemos obtenido deteniendo el reloj durante este periodo. Pero lo sí que hemos podido comprobar es el amplio eco que ha tenido en la población en general, el disponer del comodín para ralentizar el avance de la IA en nuestras vidas, esgrimiendo argumentos como el hecho de estar más organizados para dar respuestas conjuntas a la altura de las circunstancias y que se empiezan a desplegar mecanismos de control sobre la posible inundación de *fake news*.

Poco a poco vamos intuyendo hacia donde puede ir esta sensación inquietante de pérdida del control, generando todo tipo de alarmas en muchos niveles de la población. Lo que sucede es que el objeto (el robot, el chatbot, Alexa, Siri, etc en sus múltiples manifestaciones) no solo puede empezar a pensar por sí mismo (metacognición) o lo que en palabras de Bostrom (2014) se conoce como *Superinteligencia*, sino que también toma decisiones que pueden dejar, en última instancia, al margen a su progenitor y esto sí que significa un cambio de paradigma en el que estamos inmersos. Ya no se tratan de meros destellos de lo que puede llegar a conseguir, por así decirlo, una determinada aplicación como puede ser el caso de singularizar todos nuestros intereses y demandas para poder retroalimentarlas constantemente, sino de toda una realidad paralela, con un potencial que pasa a ser transhumano, donde se llega a plantear incluso el sentido de nuestra propia existencia tal y como la hemos vivido hasta ahora.

Nuestra inteligencia se está visualizando un tanto vulnerable, lenta y limitada delante de la multitud de aplicaciones de la IA. Esta es la crónica de una supervivencia anunciada y que además está siendo retransmitida cada día y en directo con las constantes propuestas de nuevas aplicaciones de la IA en cualquier terreno profesional, de hecho, en algunos momentos ya casi perdemos la cuenta, de lo que se realiza mediante la utilización de la IA y de lo que no, a pesar de que a nosotros no nos lo parezca por una profunda ausencia de transparencia. Esta ausencia genera sensaciones de desafección e inquietud. En otras palabras, la pérdida de control

contribuye a una vulnerabilidad intelectual seguido de una inseguridad donde poder refrendar la autenticidad de los datos. El episodio en el que Deep Blue derrotó en seis partidas de ajedrez a Garry Kasparov, ahora hace ya casi tres décadas, fue uno de los primeros ejemplos de esta vulnerabilidad retransmitida en directo (Campbell, Hoane and Hsu, 2002).

Delante de cualquier amenaza el ser humano reacciona de diversas formas para poder preservar su supervivencia, una de ellas es recuperando el control de la situación siendo este uno de los temas recurrentes que podemos leer en la actualidad sobre la IA. La humanidad debe regular sus políticas de desarrollo, proponiendo mecanismos de transparencia y detectar episodios de vigilancia masiva que afectan a muchos aspectos de nuestro quehacer diario (Lyon, 2001, Zuboff, 2019) por la sobreexposición de los datos que nos hacen únicos (y humanos) como puede ser el reconocimiento facial (Introna, 2025), del iris o de la huella dactilar ya implementado en nuestros ordenadores, dispositivos móviles, datos que son incluso más difíciles de conseguir (o me atrevo a decir de hackear) que nuestras contraseñas, número de identificación personal o número de la seguridad social. Todo ello siendo testigo de casos de malas praxis como son los mecanismos de suplantación de identidades, a veces, afectando a los más vulnerables como puede ser la población de mayores. Como víctimas potenciales de este agravio, uno se remite al amparo de los organismos oficiales (nacionales o internacionales) para poder regular, y si es el caso, librar la oportuna batalla para reclamar esta usurpación, siendo plenamente conscientes que ello nos sumerge en un recorrido arduo y lento.

Este amparo por parte de los organismos oficiales se está haciendo cada vez más exigente ya que también afecta a nuestro consumo de información (y posterior conocimiento) donde la nueva variable es ver si esta información es real y/o válida. El caso de las *fakenews* nos empezó a poner en sobre aviso. Ahora asistimos a la generación de las *deepfake*, vehiculado por la rapidez de acción de la IA. En este sentido se reclama la utilización de un nuevo concepto que viene a resituar la necesidad del llamado *data dignity* (Lanier & Weyl, 2018) que no es otra cosa que el trasladado del concepto dignidad humana que surgió en defensa de los totalitarismos y contraponerlo a un contexto contemporáneo, donde nuestros datos precisan de una protección mayor delante de las *nuevas* concentraciones de poder. También Han (2017) nos avisa de las múltiples formas de control como es el control social. Según este autor, el control social nos lleva a la manipulación de las emociones y conducta humanas. Por otro lado, la IA nos imprime que los algoritmos sean usados para hacer predicciones, que como nos plantea Eubanks (2018) pueden ser tan abusivas y estremecedoras como elaborar modelos estadísticos para poder predecir qué niños podrán ser posibles víctimas de abuso. Akst (2023) nos plantea en su artículo si la IA debe tener derechos legales y morales en el momento de desarrollar algunas tareas sensibles que actualmente están a cargo de los humanos, como puede ser el cuidado de mayores.

Todo ello permeabiliza la formación inicial de las futuras generaciones de profesionales. La necesidad imperiosa de tener que reconducir hacia otros niveles que no puedan ser sobrepasados, de momento, por la IA, buscando en elementos como son las llamadas habilidades blandas (*softskills*). Estas habilidades que resultan en la mejora de las interacciones entre las personas serán más que nunca, la mejor versión de nosotros mismos delante de este contexto de la IA, es decir, el poder trabajar en mayor o menor profundidad e incidir en: la comunicación, la resolución de problemas, el trabajo en equipo, la capacidad de liderazgo, la empatía, la creatividad y la adaptabilidad. Así, podremos proponer momentos de convivencia

entre lo que nos puede ofrecer el uso de los LLMs en nuestras actividades académicas, previa comprobación de su veracidad en el volcado de la información solicitada. La educación transformadora puede beneficiarse de la promoción y el uso de las softskills para mejorar el emparejamiento entre la persona y el algoritmo.

## REFERENCIAS BIBLIOGRÁFICAS

- Akst, D. (2023). Should robots with artificial intelligence have moral or legal rights? *The Wall Street Journal*.
- Arendt, H. (1951; 2017) *The origins of totalitarianism*. (e-book). Penguin Classics.
- Bostrom, N. (2014). *Superintelligence: paths, dangers, strategies*. Oxford University Press.
- Campbell, M. Hoane, A.J. y Hsu, F-h. (2002). Deep Blue. *Artificial Intelligence*, 134 (1-2), 57–83.
- Eubanks, V. (2018). *Automated Inequality: How high-tech tools profile, police and punish the poor*. St. Martin's Press.
- Han, B. (2017). *Psychopolitics: Neoliberalism and the new technologies of power*. Verso Books.
- Introna, L. (2005). Disclosive ethics and Information Technology: disclosing facial recognition systems, *Ethics and Information Technology*, 7, 75-86.
- Lanier, J. y Weyl, E. (2018). A Blueprint for a Better Digital Society. For individuals and platforms, the future requires a fundamental economic shift. *Harvard Business Review*.
- Lee, A. (2024). Staying ahead with generative artificial intelligence for learning: challenges and opportunities. *Asia Pacific Journal of Education*.  
<https://doi.org/10.1080/02188791.2024.2305171> (published online first)
- Lyon, D. (2001). *The surveillance society: monitoring everyday life*. McGraw-Hill Education. Open University Press
- Zuboff, S. (2019). *The age of surveillance capitalism: the fight for the future at the new frontier of power*. Profile Books.